

Optimal Iris Fuzzy Sketches

J. Bringer, H. Chabanne, G. Cohen, B. Kindarji and G. Zémor[‡]

Abstract

Fuzzy sketches, introduced as a link between biometry and cryptography, are a way of handling biometric data matching as an error correction issue. We focus here on iris biometrics and look for the best error-correcting code in that respect. We show that two-dimensional iterative min-sum decoding leads to results near the theoretical limits. In particular, we experiment our techniques on the Iris Challenge Evaluation (ICE) database and validate our findings.

Keywords. Iris, biometry, fuzzy sketches, min-sum decoding.

1 Introduction

Fuzzy Sketches have been introduced to handle differences occurring between two captures of biometric data, viewed as errors over a codeword. Many papers give applications of this technique for cryptographic purposes [1, 3, 2, 6, 7, 9, 17] but only a few investigate what are the best codes for this decoding problem, e.g. [8], and how to find them. This issue is addressed here.

1.1 Biometric matching and errors correction

Typically, a biometric-based recognition scheme consists of two phases: The enrollment phase where a biometric template b is measured from a user U and then registered in a token or a database. The second phase – the verification – captures a new biometric sample b' from U and compares it to the reference data via a matching function. According to some underlying measure μ and some recognition threshold τ , b' will be accepted as a biometric measure of U if $\mu(b, b') \leq \tau$, else rejected. Mainly two kinds of errors are associated to this scheme: False Reject (**FR**), when a matching user, i.e. a legitimate user, is rejected; False Acceptance (**FA**), when a non-matching one, e.g. an impostor, is accepted.

Note that, when the threshold increases, the **FR**'s rate (**FRR**) decreases while the **FA**'s rate (**FAR**) grows, and conversely.

Our methods will resort to information theory and coding. For more background, notation and classical results, the reader is referred to [4] and [11] in these two fields respectively.

Assuming that the templates live in the Hamming space $\mathcal{H} = \{0, 1\}^n$ equipped with the Hamming distance $d_{\mathcal{H}}$, the main idea of fuzzy sketches, as introduced in [9], is to convert the matching step into an error-correcting one. Let C be an error-correcting code included in \mathcal{H} :

^{*}J. Bringer and H. Chabanne are with Sagem Défense Sécurité, Eragny, France.

[†]G. Cohen is with ENST, Département Informatique et Réseaux, Paris, France.

[‡]G. Zémor is with Institut de Mathématiques de Bordeaux, Université de Bordeaux I, Bordeaux, France.

- During the enrollment phase, one stores $z = c \oplus b$, where c is a random codeword in C ,
- During the verification phase, one tries to correct the corrupted codeword $z \oplus b' = c \oplus (b \oplus b')$. Note that when the Hamming distance $d_{\mathcal{H}}(b, b')$ is small, recovering c from $c \oplus (b \oplus b')$ is, in principle, possible.

The correction capacity of C may thus be equal to τ if we do not want to alter the **FRR** and the **FAR** of the system. Unfortunately, the difference between two measures of one biometric source can be very important, whereas the correction capacity of a code is structurally constrained. Moreover, the size of the code should not be too small, to prevent z from revealing too much information about the template b .

1.2 Organization of this work

In a first part, we look for theoretical limits. We first modelize our problem with a binary erasure-and-error channel. Given a database of biometric data, we then give a method for finding an upper bound on the underlying error correction capacity.

In a second part, restricting ourselves to iris biometric data and illustrating our method with iterative min-sum decoding of product codes, we provide parameters that put our performances close to the theoretical limit.

2 Model

We consider two separate channels with a noise model based on the differences between any two biometric templates.

- The first channel, called the **matching channel**, is generated by errors $b \oplus b'$ where b and b' come from the same user U .
- The second channel, the **non-matching channel**, is generated by errors where b and b' come from different biometric sources.

In a practical biometric system, the number of errors in the **matching channel** is on average lower than in the **non-matching channel**.

Moreover, the templates are not restricted to a constant length. Indeed, when a sensor captures biometric data, we want to keep the maximum quantity of information but it is rarely possible to capture the same amount of data twice – for instance an iris may be occulted by eyelids – hence the templates are of variable length. This variability can be smoothed by forming a list of erasures, i.e. the list of coordinates where they occur. More precisely, in coding theory, an erasure in the received message is an unknown symbol at a known location. We thus have an erasure-and-error decoding problem on the **matching channel**. Simultaneously, to keep the **FAR** low, we want a decoding success to be unlikely on the **non-matching channel** : to this end we impose bounds on the correction capacity.

In the sequel, we deal with binary templates with at most N bits and assume, for the theoretical analysis that follows, that the probabilities of error and erasure on each bit are independent. Note that resorting to interleaving makes this hypothesis valid for all practical purposes.

2.1 Theoretical limit

Our goal is to estimate the capacity, in the Shannon sense [15], of the matching channel when we work with a code of a given dimension. Namely, we want to know the maximum number of errors and erasures between two biometric measures that we can manage with fuzzy sketches for this code.

Starting with a representative range of matching biometric data, the theorem below gives an easy way to estimate the lowest achievable **FRR**. The idea is to check whether the best possible code with the best generic decoding algorithm, i.e. a **maximum-likelihood (ML)** decoding algorithm (which systematically outputs the most likely codeword), would succeed in correcting the errors.

Theorem 1 *Let $k \in \mathbb{N}^*$, C be a binary code of length N and size 2^k , and m a random received message, from a random codeword of C , of length N with w_n errors and w_e erasures. Assume that C is an optimal code with respect to N and k , equipped with an **ML** decoder.*

If $\frac{w_n}{N-w_e} > \theta$ then m is only decodable with a negligible probability, where θ is such that the Hamming sphere of radius $(N - w_e)\theta$ in $\mathbb{F}_2^{N-w_e}$ contains 2^{N-w_e-k} elements.

Proof. In the case of errors only (i.e. no erasures) with error-rate $p := w_e/N$, the canonical second theorem of Shannon asserts that there are families of codes with (transmission) rate $R := k/n$ coming arbitrarily close to the channel capacity $\kappa(p)$, decodable with ML-decoding and a vanishing (in N) word error probability P_e .

In this case, $\kappa(p) = 1 - h(p)$, where $h(p)$ is the (binary) entropy function (log's are to the base 2):

$$h(x) = -x \log x - (1 - x) \log(1 - x).$$

Furthermore, P_e displays a threshold phenomenon: for any rate arbitrarily close to, but above capacity and any family of codes, P_e tends to 1 when N grows.

Equivalently, given R , there exists an error-rate threshold of

$$p = h^{-1}(1 - R),$$

h^{-1} being the inverse of the entropy function.

Back to the errors-and-erasures setting now. Our problem is to decode to the codeword nearest to the received word on the *nonerased* positions.

Thus we are now faced with a punctured code with length $N - w_e$, size 2^k , transmission rate $R' := k/(N - w_e)$ and required to sustain an error-rate $p' := \frac{w_n}{N-w_e}$.

By the previous discussion, if

$$p' > \theta := h^{-1}(1 - R'),$$

NO code and NO decoding procedure exist with a non-vanishing probability of success.

To conclude the proof, use the classical Stirling approximation for the size of a Hamming sphere of radius αM in \mathbb{F}_2^M by $2^{h(\alpha M)}$. ■

Practical implications of this theorem are illustrated in Table 1, Sec. 3.3.

3 Application

3.1 Description of the two-dimensional iterative min-sum decoding algorithm

A binary *linear* error-correcting code C is a vector subspace of \mathbb{F}_2^N . The minimum distance d_{min} of C is the smallest Hamming distance between two distinct codewords. When k is the dimension of the subspace C , i.e. when it contains 2^k codewords, C is denoted by $[N, k, d_{min}]_2$. The correction capacity t of C is the radius of the largest Hamming ball for which, for any $x \in \mathbb{F}_2^N$, there is at most one codeword in the ball of radius t centered on x . Clearly, $t = \lfloor (d_{min} - 1)/2 \rfloor$. An altered codeword with w_n errors and w_e erasures can always be corrected (by ML decoding) provided $2w_n + w_e < d_{min}$. However, if the code admits an iterative decoding algorithm, practical results overtake this limitation.

We will work with product codes together with a specific iterative decoding algorithm described below. A product code $C = C_1 \otimes C_2$ is constructed from two codes: $C_1[N_1, k_1, d_1]_2$ and $C_2[N_2, k_2, d_2]_2$. The codewords of C can be viewed as matrices of size $N_2 \times N_1$ whose rows are codewords of C_1 and columns are codewords of C_2 . This yields a $[N_1 \times N_2, k_1 \times k_2, d_1 \times d_2]$ code. When k_1 and k_2 are small enough for C_1 and C_2 to be decoded exhaustively a very efficient iterative decoding algorithm is available, namely the *min-sum* decoding algorithm. Min-sum decoding of LDPC codes was developed by Wiberg [18] as a particular instance of message passing algorithms. In a somewhat different setting it was also proposed by Tanner [16] for decoding generalized LDPC (Tanner) codes. The variant we will be using is close to Tanner's algorithm and is adapted to product codes. Min-sum is usually considered to perform slightly worse than the more classical sum-product message passing algorithm on the Gaussian, or binary-symmetric channels, but it is specially adapted to our case where knowledge of the channel is poor, and the emphasis is simply to use the Hamming distance as the appropriate basic cost function.

Let (x_{ij}) be a vector of $\{0, 1\}^{N_1 \times N_2}$. The min-sum algorithm associates to every coordinate x_{ij} a cost function κ_{ij} for every iteration of the algorithm. The cost functions are defined on the set $\{0, 1\}$. The initial cost function κ_{ij}^0 is defined by $\kappa_{ij}^0(x) = 0$ if the received symbol on coordinate (ij) is x and $\kappa_{ij}^0(x) = 1$ if the received symbol is $1 - x$.

A *row* iteration of the algorithm takes an *input* cost function κ_{ij}^{in} and produces an *output* cost function κ_{ij}^{out} . The algorithm first computes, for every row i and for every codeword $c = (c_1 \dots c_{N_1})$ of C_1 , the *sum*

$$\kappa_i(c) = \sum_{j=1}^{N_1} \kappa_{ij}^{in}(c_j)$$

which should be understood as the cost of putting codeword c on row i . The algorithm then computes, for every i, j , κ_{ij}^{out} defined as the following *min*, over the set of codewords of C_1 ,

$$\kappa_{ij}^{out}(x) = \min_{c \in C_1, c_j=x} \kappa_i(c).$$

This last quantity should be thought of as the minimum cost of putting the symbol x on coordinate (ij) while satisfying the row constraint.

A *column* iteration of the algorithm is analogous to a row iteration, with simply the roles

of the row and column indexes reversed, and code C_2 replacing code C_1 . Precisely we have

$$\kappa_j(c) = \sum_{i=1}^{N_2} \kappa_{ij}^{in}(c_i)$$

and

$$\kappa_{ij}^{out}(x) = \min_{c \in C_2, c_i=x} \kappa_j(c).$$

The algorithm alternates row and column iterations as illustrated by Fig. 1. After a given number of iterations (or before, if we find a codeword) it stops, and the value of every symbol x_{ij} is put at $x_{ij} = x$ if $\kappa_{ij}^{out}(x) < \kappa_{ij}^{out}(1 - x)$. If $\kappa_{ij}^{out}(x) = \kappa_{ij}^{out}(1 - x)$ then the value of x_{ij} stays undecided (or erased).

The following theorem is fairly straightforward to prove and illustrates the power of min-sum decoding.

Theorem 2 *If the number of errors is less than $d_1 d_2 / 2$, then two iterations of min-sum decoding of the product code $C_1 \otimes C_2$ recover the correct codeword.*

3.2 Our setting

To validate our approach, we now present the results of experiments on a practical iris database where we obtain correction performances close to the theoretical limit.

The database used for these experiments is the ICE (Phase I) database [10, 13] which contained 2953 images from 244 different eyes. A 256-byte (2048 bits) iris template, together with a 256-byte mask, is computed from each iris image using the algorithm reported in [5]; the mask filters out the unreliable bits, i.e. stores the erasures indices of the iris template. The database is taken without any modification but two slight corrections: one eye is suppressed due to a very low quality and the side of another eye has been switched from left to right. Hence we keep 2952 images. Note that in the database, the number of images provided for each eye is variable: so the number of intra-eye matching verifications between two iris codes from the same eye is not constant. The same holds for the inter-eye matching between two iris codes from different eyes. Among all the combinations, it gives a set of 29827 intra-eye matching and about 4 million of inter-eye matching to check.

The classical way to compare two iris codes I_1, I_2 with masks M_1, M_2 is to compute the relative Hamming distance

$$\frac{||(I_1 \oplus I_2) \cap M_1 \cap M_2||}{||M_1 \cap M_2||}$$

for some rotations of the second template – to deal with the iris orientation’s variation – and to keep the lowest score. It gives the following distributions of matching scores (cf. Fig 2) where we see an overlap between the two curves. We also see that the number of errors to handle in the matching channel is large (for instance at least 29% of errors for a **FRR** lower than 5%). On this channel, an additional difficulty originates from the number of erasures which varies from 512 to 1977.

3.3 Results on ICE database

We have experimented with the algorithm described in section 3.1 on this database with a particular choice for the code. In fact, the product code is constructed to fit with an array of

$$\begin{array}{c}
i \left(\begin{array}{c} \vdots \\ \hline \kappa_{i1}^{in} \quad \cdots \quad \kappa_{iN_1}^{in} \\ \hline \vdots \end{array} \right) \\
\Downarrow \\
\kappa_{ij}^{out}(x) = \min_{c \in C_1, c_j = x} \sum_{k=1}^{N_1} \kappa_{ik}^{in}(c_k) \\
\Downarrow \\
i \left(\begin{array}{c} \vdots \\ \hline \cdots \quad \kappa_{ij}^{out} \quad \cdots \\ \hline \vdots \end{array} \right) \\
\Downarrow \\
j \\
\left(\begin{array}{c} \left| \begin{array}{c} \kappa_{1j}^{in} \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \kappa_{N_2j}^{in} \end{array} \right| \\ \vdots \\ \left| \begin{array}{c} \kappa_{1j}^{in} \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \kappa_{N_2j}^{in} \end{array} \right| \end{array} \right) \\
\Downarrow \\
\kappa_{ij}^{out}(x) = \min_{c \in C_2, c_i = x} \sum_{l=1}^{N_2} \kappa_{lj}^{in}(c_l) \\
\Downarrow \\
j \\
\left(\begin{array}{c} \left| \begin{array}{c} \vdots \\ \vdots \\ \vdots \\ \vdots \\ \vdots \end{array} \right| \\ \vdots \\ \left| \begin{array}{c} \vdots \\ \vdots \\ \vdots \\ \vdots \\ \vdots \end{array} \right| \end{array} \right)
\end{array}$$

Figure 1: A row iteration followed by a column one

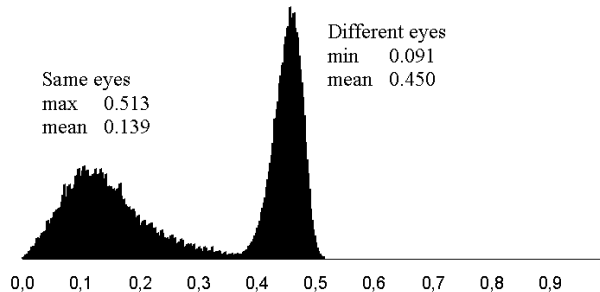


Figure 2: Inter-eyes and intra-eye distributions

2048 bits, by using Reed-Muller codes [12, 14] of order 1 which are known to have good weight distributions. A binary Reed-Muller code of order 1 in m variables, abbreviated as $RM(1, m)$, is a $[2^m, m + 1, 2^{m-1}]_2$ code. We chose to combine the $RM(1, 6)$ with the $RM(1, 5)$, leading to a product code of dimension 42 and codewords of length 64×32 .

As the density of errors and erasures in an iris code can be very high in some regions, we also added a randomly chosen interleaver to break this structure and increase the efficiency of the decoding algorithm. In so doing, we succeeded in obtaining a **FRR** of about 5.62% for a very small **FAR** (lower than 10^{-5}). This is in fact very close to the **FAR** obtained in a classical matching configuration for a similar **FRR**.

The overall size of the code could appear small from a cryptographic point of view, but following the theoretical analysis of section 2.1, it is difficult to expect much more while achieving a low **FRR** on this database. Indeed, from the distribution of errors and erasures on the **matching channel**, we obtain by Theorem 1 the practical limits which are reported in Table 1.

Table 1: Theoretical limits on ICE database

Code's dimension	Best theoretical FRR
42	2.49%
64	3.76%
80	4.87%
128	9.10%

Remark. In [8], the fuzzy sketch scheme is applied with a concatenated error-correcting code combining a Hadamard code and a Reed-Solomon code. More precisely, the authors use a Reed-Solomon code of length 32 over \mathbb{F}_{27} (with a correction capacity $t_{RS} < 16$) and a Hadamard code of order 6 and length 64 (with a correction capacity $t_H = 15$): a codeword of 2048 bits is in fact constructed as a set of 32 blocks of 64 bits where each block is a codeword of the underlying Hadamard code. As explained in [8], the Hadamard code is introduced to deal with the background errors and the Reed-Solomon code to deal with the bursts (e.g. caused by eyelashes, reflections, ...).

Note that in this scheme, the model is not exactly the same as ours, as the masks are not taken into account. Moreover, the quality of the database used in [8] is better than for the ICE database. Actually, [8] reports very good results on their experiments with a database

of 700 images, but the codes do not seem appropriate to our case as our experiment on the ICE database gave a too large rate of **FR** (e.g. 10% of **FR** with 0.80% of **FA**), even for the smallest possible dimension of the Reed-Solomon code when $t_{RS} = 15$.

4 Conclusion

We derived explicit upper bounds on the correction capacity of Fuzzy Sketches on iris-based biometrics. We then showed how the two-dimensional iterative min-sum decoding algorithm achieves correction performance close to the optimal decoding rate. Our results were validated on a typical iris database.

References

- [1] X. Boyen, Y. Dodis, J. Katz, R. Ostrovsky, and A. Smith. Secure remote authentication using biometric data. In R. Cramer, editor, *EUROCRYPT*, volume 3494 of *Lecture Notes in Computer Science*, pages 147–163. Springer, 2005.
- [2] G. Cohen and G. Zémor. Generalized coset schemes for the wire-tap channel: application to biometrics. In *ISIT*, 2004.
- [3] G. Cohen and G. Zémor. The wire-tap channel applied to biometrics. In *ISITA*, 2004.
- [4] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. Wiley, 2006.
- [5] J. Daugman. The importance of being random: statistical principles of iris recognition. *Pattern Recognition*, 36(2):279–291, 2003.
- [6] G. I. Davida and Y. Frankel. Perfectly secure authorization and passive identification for an error tolerant biometric system. In M. Walker, editor, *IMA Int. Conf.*, volume 1746 of *Lecture Notes in Computer Science*, pages 104–113. Springer, 1999.
- [7] Y. Dodis, L. Reyzin, and A. Smith. Fuzzy extractors: How to generate strong keys from biometrics and other noisy data. In C. Cachin and J. Camenisch, editors, *EUROCRYPT*, volume 3027 of *Lecture Notes in Computer Science*, pages 523–540. Springer, 2004.
- [8] F. Hao, R. Anderson, and J. Daugman. Combining crypto with biometrics effectively. *IEEE Transactions on Computers*, 55(9):1081–1088, 2006.
- [9] A. Juels and M. Wattenberg. A fuzzy commitment scheme. In *ACM Conference on Computer and Communications Security*, pages 28–36, 1999.
- [10] Xiaomei Liu, Kevin W. Bowyer, and Patrick J. Flynn. Iris Recognition and Verification Experiments with Improved Segmentation Method. In *Fourth IEEE Workshop on Automatic Identification Advanced Technologies (AutoID)*, 17-18 October 2005, Buffalo, New York, 2005.
- [11] F. J. MacWilliams and N. J. A. Sloane. *The Theory of Error-correcting Codes*. North-Holland, 1988.

- [12] D.E. Muller. Application of boolean algebra to switching circuit design and to error detection. *IEEE Trans. on Electronic Computers*, 3:6–12, 1954.
- [13] National Institute of Science and Technology (NIST). Iris Challenge Evaluation. <http://iris.nist.gov/ICE>, 2005.
- [14] I.S. Reed. A class of multiple-error-correcting codes and their decoding scheme. *IEEE Trans. on Information Theory*, 4:38–42, 1954.
- [15] Claude E. Shannon. A mathematical theory of communication. *The Bell System Technical Journal*, 27, 1948. <http://cm.bell-labs.com/cm/ms/what/shannonday/paper.html>.
- [16] R. M. Tanner. A recursive approach to low-complexity codes. *IEEE Trans. on Information Theory*, 27:533–547, 1981.
- [17] P. Tuyts and J. Goseling. Capacity and examples of template-protecting biometric authentication systems. In D. Maltoni and A. K. Jain, editors, *ECCV Workshop BioAW*, volume 3087 of *Lecture Notes in Computer Science*, pages 158–170. Springer, 2004.
- [18] N. Wiberg. *Codes and Decoding on general Graphs*. PhD thesis, Linköping University, Linköping, Sweden, 1996.